"All the News That's Fit to Print"

The New York Times

National Edition
Northern California: Mostly sunny, warm and breezy at times. Highs in 70s and 80s, but 60s in mountains. Tomorrow, sunny and warm. Weather map, Page D10.

VOL. CLVII . . . No. 54,105
© 2007 The New York Times
MONDAY, OCTOBER 22, 2007

Printed in California $1.25

INTERNET ARCHIVE

Brewster Kahle, Founder and Digital Librarian
brewster@archive.org, 415.561.6767

Robert Miller, Director of Books
robert@archive.org, 415.640.1092

LIBRARIES SHUN DEALS TO PLACE BOOKS ON WEE

TECH GIANTS REBUFFEI

Scanning by Microsoft and Google Comes With Restrictions

By KATIE HAFNER

Several major research libraries have rebuffed offers from Google and Microsoft to scan their books into computer databases, saying they are put off by restrictions these companies want to place on the new digital collections.

The research libraries, including a large consortium in the Boston area, are instead signing on with the Open Content Alliance, a nonprofit effort aimed at making their materials broadly available.

Libraries that agree to work with Google must agree to a set of terms, which include making the material unavailable to other commercial search services. Microsoft places a similar restriction on the books it converts to electronic form. The Open Content Alliance, by contrast, is making the material available to any search service.

Google pays to scan the books and does not directly profit from the resulting Web pages, although the books make its search engine more useful and more valuable. The libraries can have their books scanned again by another company or organization for dissemination more broadly.

It costs the Open Content Alliance as much as $30 to scan each book, a cost shared by the group's members and benefactors, so there are obvious financial benefits to libraries of Google's wide-ranging offer, started in 2004.

Many prominent libraries have accepted Google's offer — including the New York Public Library and libraries at the University of Michigan, Harvard, Stanford and Oxford. Google expects to scan 15 million books from those collections over the next decade.

But the resistance from some libraries, like the Boston Public

Continued on Page A18

Libraries Rejecting Deals To Scan Books for the Web

From Page A1

Library and the Smithsonian Institution, suggests that many in the academic and nonprofit world are intent on pursuing a vision of the Web as a global repository of knowledge that is free of business interests or restrictions.

Even though Google's program could make millions of books available to hundreds of millions of Internet users for the first time, some libraries and researchers worry that if any one company comes to dominate the digital conversion of these works, it could exploit that dominance for commercial gain.

"There are two opposed pathways being mapped out," said Paul Duguid, an adjunct professor at the School of Information at the University of California, Berkeley. "One is shaped by commercial concerns, the other by a commitment to openness, and which one will win is not clear."

Last month, the Boston Library Consortium of 19 research and academic libraries in New England that includes the University of Connecticut and the University of Massachusetts, said it would work with the Open Content Alliance to begin digitizing the books among the libraries' 34 million volumes whose copyright had expired.

"We understand the commercial value of what Google is doing, but we want to be able to distribute materials in a way where everyone benefits from it," said Bernard A. Margolis, president of the Boston Public Library, which has in its collection roughly 3,700 volumes from the personal library of John Adams.

Mr. Margolis said his library had spoken with both Google and Microsoft, and had not shut the door entirely on the idea of working with them. And several libraries are working with both Google and the Open Content Alliance.

Adam Smith, project management director of Google Book Search, noted that the company's deals with libraries were not exclusive. "We're excited that the O.C.A. has signed more libraries, and we hope they sign many more," Mr. Smith said.

"The powerful motivation is that we're bringing more offline information online," he said. "As a commercial company, we have the resources to do this, and we're doing it in a way that benefits users, publishers, authors and libraries. And it benefits us because we provide an improved user experience, which then means users will come back to Google."

The Library of Congress has a pilot program with Google to digitize some books. But in January, it announced a project with a more inclusive approach. With $2 million from the Alfred P. Sloan Foundation, the library's first mass digitization effort will make 136,000 books accessible to any search engine through the Open Content Alliance. The library declined to comment on its future digitization plans.

The Open Content Alliance is the brainchild of Brewster Kahle, the founder and director of the Internet Archive, which was created in 1996 with the aim of preserving copies of Web sites and other material. The group includes more than 80 libraries and research institutions, including the Smithsonian Institution.

Although Google is making public-domain books readily available to individuals who wish to download them, Mr. Kahle and others worry about the possible implications of having one company store and distribute so much public-domain content.

"Scanning the great libraries is a wonderful idea, but if only one corporation controls access to this digital collection, we'll have handed too much control to a private entity," Mr. Kahle said.

The Open Content Alliance, he said, "is fundamentally different, coming from a community project to build joint collections that can be used by everyone in different ways."

Mr. Kahle's group focuses on out-of-copyright books, mostly those published in 1922 or earlier. Google scans copyrighted works as well, but it does not allow users to read the full text of those books online, and it allows publishers to opt out of the program.

Microsoft joined the Open Content Alliance at its start in 2005, as did Yahoo, which also has a book search project. Google also spoke with Mr. Kahle about joining the group, but they did not reach an agreement.

A year after joining, Microsoft added a restriction that prohibits a book it has digitized from being included in commercial search engines other than Microsoft's.

"Unlike Google, there are no restrictions on the distribution of these copies for academic purposes across institutions," said Jay Girotto, group program manager for Live Book Search from Microsoft. Institutions working with Microsoft, he said, include the University of California and the New York Public Library.

Some in the research field view the issue as a matter of principle.

Doron Weber, a program director at the Sloan Foundation, which has made several grants to libraries for digital conversion of books, said that several institutions approached by Google have spoken to his organization about their reservations. "Many are hedging their bets," he said, "taking Google money for now while realizing this is, at best, a short-term bridge to a truly open universal library of the future."

The University of Michigan, a Google partner since 2004, does not seem to share this view. "We have not felt particularly restricted by our agreement with Google," said Jack Bernard, a lawyer at the university.

The University of California, which started scanning books with the Open Content Alliance, Microsoft and Yahoo in 2005, has added Google. Robin Chandler, director of data acquisitions at the University of California's digital library project, said working with everyone helps increase the volume of the scanning.

Some have found Google to be inflexible in its terms. Tom Garnett, director of the Biodiversity Heritage Library, a group of 10 prominent natural history and botanical libraries that have agreed to digitize their collections, said he had had discussions with various people at both Google and Microsoft.

"Google had a very restrictive agreement, and in all our discussions they were unwilling to yield," he said. Among the terms was a requirement that libraries put their own technology in place to block commercial search services other than Google, he said.

Libraries that sign with the Open Content Alliance are obligated to pay the cost of scanning the books. Several have received grants from organizations like the Sloan Foundation.

The Boston Library Consortium's project is self-funded, with $845,000 for the next two years. The consortium pays 10 cents a page to the Internet Archive, which has installed 10 scanners at the Boston Public Library.Other members include the Massachusetts Institute of Technology and Brown University.

The scans are stored at the Internet Archive in San Francisco and are available through its Web site. Search companies including Google are free to point users to the material.

On Wednesday the Internet Archive announced, together with the Boston Public Library and the library of the Marine Biological Laboratory and Woods Hole Oceanographic Institution, that it would start scanning out-of-print but in-copyright works to be distributed through a digital interlibrary loan system.

ANN JOHANSSON FOR THE NEW YORK TIMES

Workers scan books for the Internet Archive. Its founder conceived the Open Content Alliance to distribute library books online.